



TITLE:

ランダム行列理論を用いた気象データ相関抽出とその有効性の検討(経済物理学とその周辺,統計数理研究所研究会共同研究集会,経済物理学2009-ミクロとマクロの架け橋-,京都大学基礎物理学研究所2009年度前期研究会,研究会報告)

AUTHOR(S):

村上, 和正; 安本, 典生; 家富, 洋

CITATION:

村上, 和正 ...[et al]. ランダム行列理論を用いた気象データ相関抽出とその有効性の検討(経済物理学とその周辺,統計数理研究所研究会共同研究集会,経済物理学2009-ミクロとマクロの架け橋-,京都大学基礎物理学研究所2009年度前期研究会,研究会報告). 物性研究 2010, 93(5): 706-707

ISSUE DATE:

2010-02-05

URL:

<http://hdl.handle.net/2433/169209>

RIGHT:

ランダム行列理論を用いた気象データ相関抽出とその有効性の検討

新潟大学大学院 自然科学研究科 村上 和正, 安本 典生
新潟大学 理学部 家富 洋

本研究の目的は, ランダム行列理論 (RMT) を用いた時系列データ解析 [1-3] によって有意な相関構造が抽出されているかを検証することである。

1 はじめに

我々は, 真の気象相関構造を明らかにするために統計的不確定性のあるデータを解析している。特に異常気象をもたらす原因の一つであるテレコネクションに着目している。しかし RMT を用いた時系列データ解析によって本当に有意な相関構造が明らかになっているのかという疑問が残る。その疑問に答えるため, 素性の良く知れた多変量正規分布モデル (ガウスモデル) を考えた。このモデルでは任意の相関行列を与えることでデータを幾らでも生成することができる。気象実データは一つの地球に対してごく限られた期間内で測定されたものであり, その解析結果がどの程度の有効性を持つものか判断することは難しい。だが, 時系列データに対応する相関を持ったサンプルをガウスモデルを使って生成し, それらに対して実データと同様の解析方法をとることによって, 与えた相関構造が抽出できているのかを検証することができる。つまりその結果から解析法が有効であるのかを確認することができる。短所としては, 各時系列データの分布が正規分布になる事と, 異時刻相関を表現できない点がある。データ数や時系列長も任意に決められるが, 今回は相関行列を実データと比較するためにデータ数を 1033, 時系列長を 1538 と合わせた。

2 サンプル相関行列の固有値, 固有ベクトル

まずサンプルデータの固有値を調べてみた。サンプルデータの最大固有値等の主要な固有値の大きさは, 実データと比べて 2 % 程度の誤差は持っていたが, 全体の分布形は変わってはいなかった。複数回検証したが分布形が変わったと判断できるようデータは表れなかった。これは統計的不確定性は, 本質的な分布の変化を引き起こさないことを示唆している。

図 1 は, 縦軸に χ^2 二乗値, 横軸に固有ベクトルの番号 (モード) j をプロットしてある。今回は, 固有ベクトル成分を $[-5, +5]$ の区間で 21 個の箱に分けた。

$$\chi_j^2 = \sum_{n=1}^{21} \frac{(f_{jn} - m_{jn})^2}{f_{jn}} \quad (1)$$

ここで f_{jn} は理論値, m_{jn} はサンプルデータにおける j 番目のモードの n 番目の箱に入るデータの数である。実データにおいては部分的に $j=200$ 付近 (図中の丸印) で大きくなる様子が現れているが, それはサンプルデータでも同様であった。また実データの大きな固有値に対応するモード (図中の矢印) に, 大きな χ^2 二乗値を持つものがあるが, サンプルデー

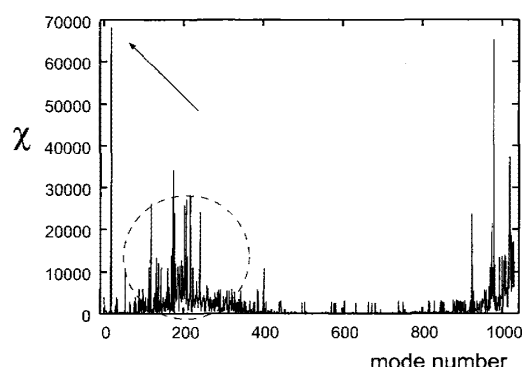


図 1: 固有ベクトル成分分布 χ 二乗値による RMT との比較

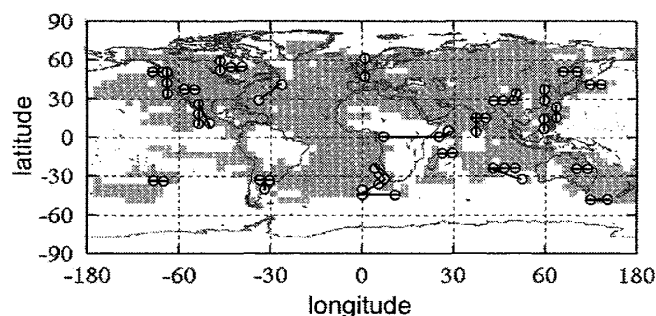


図 2: 地球表面上に表示したノイズ相関行列の影響

タによっては出現したりしなかったりする。確率として 100 回試行してみたところ, 50 番目までのモードに対して χ 二乗値 20000 以上の値を持つ確率は, 20 % である事が分かった。この確率は, 十分に有意に発生する確率であると言える。つまり実データにおける異常と判断されそうなモードも無視できない。よって, ガウスモデルによって作られたサンプルデータは, 理論値とのずれ方に統計的不確定性は持つが, 全体的な特徴は再現していると結論できる。

3 主成分相関行列

相関行列を λ_+ (RMT で与えられる最大固有値) 以上のスペクトルで構成される主成分相関行列とそれ以外のノイズ相関行列に分ける [3]。もし, 主成分相関行列によって十分な相関が抽出できているのならば, ノイズ相関行列の寄与は小さく無視できると考えられる。具体的に数値計算を実行すると, ノイズ相関行列は, 最大で 0.46 の無視できない寄与を持つことが分かった。そのような大きな寄与が空間的にどのような地点に対応しているかを図示したのが図 2 である。0.2 以上の相関係数の差異を持つデータ点を○で描き, 相関する地点同士を線で結んだ。地球表面上で一萬キロ以上のように大きく距離が離れている点がノイズ相関行列には存在していないことが分かる。つまり無視したノイズ項は我々が着目するテレコネクションには大きな影響を与えない事が分かった。

4 まとめ

ガウスモデルの範囲内であるが, RMT を用いた時系列データ解析の有効性を確かめることができた。また χ 二乗値を見ることによって統計的不確定性の定量的評価と異常値と判断されそうなデータに対しての有意性の検証をすることができた。

- [1] L. Laloux, P. Cizeau, J.-P. Bouchaud, and M. Potters, Phys. Rev. Lett. **83**, 1467 (1999).
- [2] V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Amaral, T. Guhr, and H. E. Stanley, Phys. Rev. E **65**, 066126 (2002).
- [3] M. S. Santhanam and P. K. Patra, Phys. Rev. E **64**, 016102 (2001)
- [4] 青山 秀明, 相馬 亘, 家富 洋, 藤原 義久, 池田 裕一 『経済物理学』 第 5 章 2008.